2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517798

Guessing a Tuple

Robert Graczyk, Amos Lapidoth and Yiming Yan
ETH Zurich, 8092 Zurich, Switzerland
Email: {graczyk, lapidoth, yan}@isi.ee.ethz.ch

***Abstract*—A single-letter expression is provided for the exponential growth rate of the least expected number of guesses required to recover all the sequences produced by correlated memoryless sources when each guess is of a single source sequence, with the source at the guesser's discretion.**

I. Introduction and Problem Statement

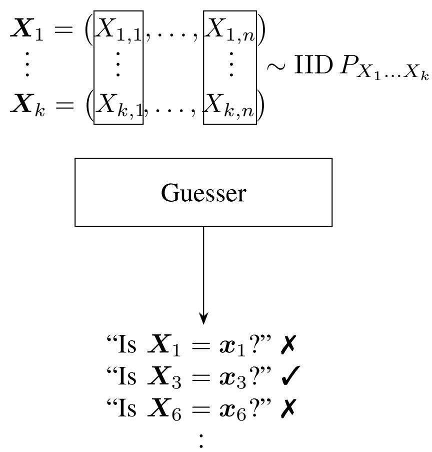

Fig. 1: Guessing the tuple $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k) \in \mathcal{X}_1^n \times \dots \times \mathcal{X}_k^n$ one component at a time.

Each of k correlated memoryless sources produces an n-length random sequence, with the j-th source producing the sequence $\boldsymbol{X}_j = (X_{j,1}, \dots, X_{j,n})$ comprising n random symbols $X_{j,1}, \dots, X_{j,n}$, each of which takes values in the finite alphabet $\mathcal{X}_j$. The n k-tuples $\{(X_{1,i}, \dots, X_{k,i})\}_{i=1}^n$ are IID according to some joint PMF $P_{X_1 \dots X_k}$ on $\mathcal{X}_1 \times \dots \times \mathcal{X}_k$:

$$\big(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k\big) \sim P^{\times n}_{X_1 \dots X_k}, \tag{1}$$

where $P^{\times n}$ denotes the n-fold product of P.

All k source sequences $\boldsymbol{X}_1, \dots, \boldsymbol{X}_k$ are to be guessed with guesses of the form

$$\text{“Is } \boldsymbol{X}_j = (\xi_1, \dots, \xi_n)\text{?”} \tag{2}$$

where the source j pertaining to the guess is at the guesser's discretion as is the n-tuple $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$, which, without loss of optimality, can be restricted to be an element of $\mathcal{X}_j^n$. To simplify notation, we shall assume that the alphabets $\mathcal{X}_1, \dots, \mathcal{X}_k$ are disjoint and that $\boldsymbol{\xi}$ is in $\mathcal{X}_1^n \cup \dots \cup \mathcal{X}_k^n$. Under this assumption, $\boldsymbol{\xi}$ specifies not only the guess but also, implicitly, which source is being guessed.

Given a guessing strategy $\mathcal{S}$, let G_l denote the number of guesses taken until the l-th affirmative answer, i.e., until l components of $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k)$ are revealed. The total number of guesses required to recover $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k)$ is therefore G_k. This can be viewed as the total number of guesses an attacker would need to recover k correlated passwords by trial and error. In Section II we characterize, for $\rho \geq 0$, the least achievable exponential growth rate of $\mathbb{E}[G_k^\rho]$ over all guessing strategies:

Theorem 1. *When $\{(X_{1,i}, \dots, X_{k,i})\}_{i=1}^n \sim \text{IID}\, P_{X_1 \dots X_k}$,*

$$\begin{aligned} \lim_{n\to\infty} \min_{\mathcal{S}} \frac{1}{n} \log \mathbb{E}[G_k^\rho] &= \sup_{Q_{X_1 \dots X_k}} \Big(\rho \min_{\pi} \max_{i} \mathrm{H}_Q\left(X_{\pi(i)} \mid X_{\pi(1)}, \dots, X_{\pi(i-1)}\right) \\ &\quad - \mathrm{D}(Q_{X_1 \dots X_k} \| P_{X_1 \dots X_k})\Big), \quad \rho \geq 0, \end{aligned} \tag{3}$$

where $\mathrm{H}_Q(\cdot)$ denotes the Shannon entropy w.r.t. $Q_{X_1 \dots X_k}$; on the left-hand side (LHS) the minimum is over all guessing strategies $\mathcal{S}$; and on the right-hand side (RHS) the supremum is over all PMFs $Q_{X_1 \dots X_k}$ on $\mathcal{X}_1 \times \dots \times \mathcal{X}_k$, and the minimum over all permutations $\pi\colon [1:k] \to [1:k]$.

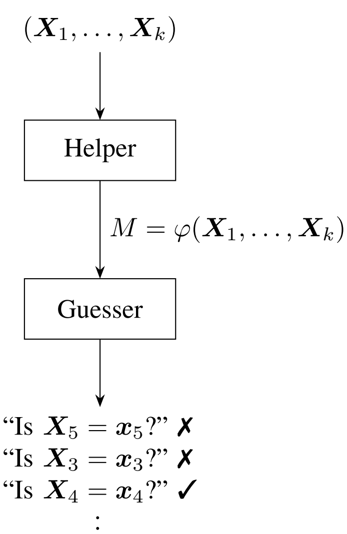

Fig. 2: Guessing $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k)$ with a helper.

Another setting we study is with a rate-R helper φ,

$$\begin{aligned} \varphi\colon\ \mathcal{X}_1^n \times \dots \times \mathcal{X}_k^n &\to \{0,1\}^{nR} \\ (\boldsymbol{X}_1, \dots, \boldsymbol{X}_k) &\mapsto M \end{aligned} \tag{4}$$

whose nR-bit description M of $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_k)$ is revealed to the guesser prior to guessing. The guesser's strategy—now depending on M—is denoted $\mathcal{S}(M)$ and the number of guesses until the l-th affirmative answer $G_l(M)$.

In Section III we characterize, for $\rho \geq 0$, the least achievable exponential growth rate of $\mathbb{E}[G_k(M)^\rho]$ over all helpers and guessing strategies:

Theorem 2. *When $\{(X_{1,i},\dots,X_{k,i})\}_{i=1}^n \sim \text{IID}\, P_{X_1\dots X_k}$, and M takes values in $\{0,1\}^{nR}$,*

$$\begin{aligned}\lim_{n\to\infty} \min_{\varphi,\mathcal{S}(M)} \frac{1}{n}\log \mathbb{E}[G_k(M)^\rho] \\ = \sup_{Q_{X_1\dots X_k}} \inf_{Q_{U|X_1\dots X_k}:\ \mathrm{I}_Q(X_1,\dots,X_k;U)\leq R} \\ \Big(\rho \min_\pi \max_i \mathrm{H}_Q\left(X_{\pi(i)} \mid X_{\pi(1)},\dots,X_{\pi(i-1)},U\right) \\ - \mathrm{D}(Q_{X_1\dots X_k}\|P_{X_1\dots X_k})\Big), \quad \rho\geq 0,\end{aligned} \tag{5}$$

where on the LHS the minimum is over all nR-bit helpers φ (as defined in (4)) and over all helper-dependent guessing strategies $\mathcal{S}(M)$; and on the RHS the supremum is over all PMFs $Q_{X_1\dots X_k}$ on $\mathcal{X}_1\times\dots\times\mathcal{X}_k$, the infimum is over the conditional PMFs $Q_{U|X_1\dots X_k}$ on $\mathcal{U}\times\mathcal{X}_1\times\dots\times\mathcal{X}_k$ satisfying the mutual information constraint (calculated w.r.t. $Q_{X_1\dots X_k}\circ Q_{U|X_1\dots X_k}$), where $\mathcal{U}$ can be any finite alphabet, and the minimum is over all permutations π on $[1:k]$.

When the sources producing $\boldsymbol{X}_1,\dots,\boldsymbol{X}_k$ are independent, our guessing problem reduces to the multi-user guesswork problem studied by Christiansen et al. [1]; when $k=1$, our guessing problem reduces to that of Massey [2] and Arıkan [3]. Other variations on the Massey-Arıkan problem include guessing with side-information [3]; guessing subject to source uncertainty [4]; guessing with a distortion criterion [5], [6]; distributed randomized guessing [7]; and guessing on the Gray-Wyner and Slepian-Wolf network [8].

II. Guessing Without a Helper

Achievability. We first prove the direct part of Theorem 1 by constructing a sequence of guessing strategies for which $\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[G_k^\rho]$ is upper-bounded by the RHS of (3). For brevity, we restrict our analysis to $\rho=1$.

We begin with some notation. For a positive integer n and a size-l subset of indices $\{i_1,\dots,i_l\}\subseteq[1:k]$, the set of all denominator-n types (rational PMFs with denominator n) on $\mathcal{X}_{i_1}\times\dots\times\mathcal{X}_{i_l}$ is denoted $\mathcal{P}^n(\mathcal{X}_{i_1}\times\dots\times\mathcal{X}_{i_l})$. The empirical distribution of a tuple $(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})\in\mathcal{X}_{i_1}^n\times\dots\times\mathcal{X}_{i_l}^n$ is denoted $\hat{Q}_{\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l}}$, so

$$\begin{aligned}&\hat{Q}_{\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l}}(x_{i_1},\dots,x_{i_l}) \\ &\quad\triangleq \frac{1}{n}\mathrm{N}(x_{i_1},\dots,x_{i_l}\mid \boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l}),\end{aligned} \tag{6}$$

where $\mathrm{N}(x_{i_1},\dots,x_{i_l}\mid \boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})$ denotes the number of occurrences of $(x_{i_1},\dots,x_{i_l})$ in $(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})$. The type class of a type $Q\in\mathcal{P}^n(\mathcal{X}_{i_1}\times\dots\times\mathcal{X}_{i_l})$ is denoted $\mathcal{T}^n(Q)$, so

$$\begin{aligned}\mathcal{T}^n(Q)\triangleq\big\{(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})\in\mathcal{X}_{i_1}^n\times\dots\times\mathcal{X}_{i_l}^n: \\ \hat{Q}_{\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l}}=Q\big\}.\end{aligned} \tag{7}$$

Given a tuple $(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})\in\mathcal{X}_{i_1}^n\times\dots\times\mathcal{X}_{i_l}^n$ and a type $\tilde{Q}\in\mathcal{P}^n(\mathcal{X}_{i_1}\times\dots\times\mathcal{X}_{i_l}\times\mathcal{X}_{i_{l+1}})$, the conditional type class of $\tilde{Q}$ given $(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})$ is denoted $\mathcal{T}^n(\tilde{Q}\mid\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l})$, so

$$\begin{aligned}&\mathcal{T}^n(\tilde{Q}\mid\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l}) \\ &\quad\triangleq\big\{\boldsymbol{x}_{i_{l+1}}\in\mathcal{X}_{i_{l+1}}^n: Q_{\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_l},\boldsymbol{x}_{i_{l+1}}}=\tilde{Q}\big\}.\end{aligned} \tag{8}$$

Finally, if $Q\in\mathcal{P}^n(\mathcal{X}_{i_1}\times\dots\times\mathcal{X}_{i_l})$, then $\mathbb{E}_Q[\cdot]$ denotes expectation with $(\boldsymbol{X}_{i_1},\dots,\boldsymbol{X}_{i_l})$ drawn equiprobably from the type class $\mathcal{T}^n(Q)$.

To prove the direct part of Theorem 1, we proceed in three steps: first, we show that w.l.o.g. the guesser can be assumed cognizant of the empirical distribution $\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}$ of $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$; second, for every n and every empirical distribution $Q\in\mathcal{P}^n(\mathcal{X}_1\times\dots\times\mathcal{X}_k)$ that $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ can assume, we construct a guessing strategy $\mathcal{S}_Q$; and third, we show that, under $\mathcal{S}_{\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}}$, the exponential growth rate of the expected number of guesses is upper-bounded by the RHS of (3).

The first step follows from an argument analogous to that in Proposition 6.9 in [9] with

$$X\leftarrow(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k),\quad Y\leftarrow\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}. \tag{9}$$

It guarantees the existence of a guessing strategy $\mathcal{S}^*$ whose expected number of guesses $\mathbb{E}[G_k^*]$ satisfies

$$\mathbb{E}[G_k^*]\leq|\mathcal{P}^n(\mathcal{X}_1\times\dots\times\mathcal{X}_k)|\cdot\min_{\mathcal{S}_\mathrm{T}}\mathbb{E}[G_k], \tag{10}$$

where the minimum on the RHS is over all type-cognizant guessing strategies $\mathcal{S}_\mathrm{T}$ (i.e., guessing strategies that may depend on $\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}$). Because the number of types $|\mathcal{P}^n(\mathcal{X}_1\times\dots\times\mathcal{X}_k)|$ is subexponential in n [9, Thm. 2.11],

$$\frac{1}{n}\log\mathbb{E}[G_k^*]\leq\min_{\mathcal{S}_\mathrm{T}}\frac{1}{n}\log\mathbb{E}[G_k]+\delta_n, \tag{11}$$

where $\{\delta_n\}$ is some suitable positive sequence that decays to zero as n tends to infinity. It thus suffices to construct a type-cognizant guessing strategy $\mathcal{S}_\mathrm{T}^*$ whose expected number of guesses grows exponentially at a rate not exceeding the RHS of (3).

To that end, we now proceed to the second step of the proof and condition on the event

$$\mathcal{A}(Q)\triangleq\{\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}=Q\}. \tag{12}$$

Because $\{(X_{1,i},\dots,X_{k,i})\}_{i=1}^n$ are IID, $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ is distributed equiprobably over $\mathcal{T}^n(Q)$ given $\mathcal{A}(Q)$,

$$\begin{aligned}&\Pr\left[(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)=(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\mid\mathcal{A}(Q)\right] \\ &\quad=\begin{cases}\frac{1}{|\mathcal{T}^n(Q)|}, & \text{if } \hat{Q}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k}=Q \\ 0, & \text{else.}\end{cases}\end{aligned} \tag{13}$$

We next fix a permutation $\pi\colon[1:k]\to[1:k]$ and propose the following guessing strategy for $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ distributed according to (13) (below, we interchangeably use $Q_{i_1,\dots,i_l}$ and $Q_{X_{i_1},\dots,X_{i_l}}$ to denote the $\{X_{i_1},\dots,X_{i_l}\}$-marginal of Q):

Guessing Strategy $\mathcal{S}_Q^\pi$

Recover $\boldsymbol{X}_{\pi(1)}$ using an arbitrary guessing order on $\mathcal{T}^n(Q_{\pi(1)})$.

for $i \leftarrow 2$ **to** k **do**

Recover $\boldsymbol{X}_{\pi(i)}$ using an arbitrary guessing order on $\mathcal{T}^n(Q_{\pi(1),\ldots,\pi(i)} \mid \boldsymbol{X}_{\pi(1)}, \ldots, \boldsymbol{X}_{\pi(i-1)})$.

end for

The strategy $\mathcal{S}_Q^\pi$ corresponds to guessing the sequences $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ one-by-one in the order determined by π exploiting the fact that

$$\boldsymbol{X}_{\pi(i)} \in \mathcal{T}^n(Q_{\pi(1),\ldots,\pi(i)} \mid \boldsymbol{X}_{\pi(1)}, \ldots, \boldsymbol{X}_{\pi(i-1)}), \tag{14}$$

which guarantees that $\boldsymbol{X}_{\pi(i)}$ is revealed after at most $|\mathcal{T}^n(Q_{\pi(1),\ldots,\pi(i)} \mid \boldsymbol{X}_{\pi(1)}, \ldots, \boldsymbol{X}_{\pi(i-1)})|$ guesses. Thus,

$$\begin{aligned}
&\mathbb{E}_Q[G_k] \\
&\quad = \sum_{i=1}^k \mathbb{E}_Q[G_i] - \mathbb{E}_Q[G_{i-1}] && (15)\\
&\quad \le \sum_{i=1}^k \mathbb{E}_Q\left[\left|\mathcal{T}^n(Q_{\pi(1),\ldots,\pi(i)} \,|\, \boldsymbol{X}_{\pi(1)}, \ldots, \boldsymbol{X}_{\pi(i-1)})\right|\right] && (16)\\
&\quad \le \sum_{i=1}^k 2^{n\mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})} && (17)\\
&\quad \le k \cdot 2^{n(\max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)}))} && (18)
\end{aligned}$$

where in (15) we have implicitly defined $G_0 \triangleq 0$; and (17) follows from [9, Thm. 2.31]. Optimizing over π leads to a guessing strategy $\mathcal{S}_Q^*$ whose expected number of guesses $\mathbb{E}_Q[G_k^*]$ satisfies

$$\mathbb{E}_Q[G_k^*] \le 2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})+\delta_n')}. \tag{19}$$

Next, we proceed to the third and final part of the proof. Define the type-cognizant guessing strategy $\mathcal{S}_\mathrm{T}^* \triangleq \mathcal{S}^*_{\hat{Q}_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}}$, namely, the strategy where the guesser observes the empirical distribution of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ (justified in Step 1) and then applies the corresponding guessing strategy constructed in Step 2. We show that, when $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k) \sim P^{\times n}_{X_1,\ldots,X_k}$, the expected number of guesses under $\mathcal{S}_\mathrm{T}^*$, namely $\mathbb{E}[G_k^*]$, grows exponentially at a rate not exceeding the RHS of (3). Indeed, averaging over $\mathcal{A}(Q)$, $Q \in \mathcal{P}^n(\mathcal{X}_1 \times \cdots \times \mathcal{X}_k)$ and using (19),

$$\begin{aligned}
&\mathbb{E}[G_k^*] \\
&\quad = \sum_Q \mathbb{E}_Q[G_k^*] \Pr[\mathcal{A}(Q)] && (20)\\
&\quad \le \sum_Q \Big(2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})+\delta_n')} \\
&\qquad\qquad \Pr[\mathcal{A}(Q)]\Big) && (21)\\
&\quad \le \sum_Q \Big(2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})+\delta_n')} \\
&\qquad\qquad 2^{-n\,\mathrm{D}(Q\|P_{X_1\ldots X_k})}\Big) && (22)\\
&\quad \le \max_Q \Big(2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})+\delta_n')} \\
&\qquad\qquad 2^{-n\,\mathrm{D}(Q\|P_{X_1\ldots X_k})}\Big) 2^{n\delta_n}, && (23)
\end{aligned}$$

where (22) follows from [9, Thm. 2.21]; and in (23) we have upper-bounded the sum by the product of the largest addend and the number of addends (number of types), with the latter's contribution to the exponential growth δ_n vanishing as $n \uparrow \infty$ (cf. (11)). Taking the limit and the logarithm on both sides of (23), and using the fact that the set of types is dense in the set of all PMFs, we conclude that for the proposed guessing strategy $\mathcal{S}_\mathrm{T}^*$

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[G_k^*] \\
&\quad \le \sup_{Q_{X_1\ldots X_k}} \Big(\min_\pi \max_i \mathrm{H}_Q\left(X_{\pi(i)} \mid X_{\pi(1)}, \ldots, X_{\pi(i-1)}\right) \\
&\qquad\qquad - \mathrm{D}(Q_{X_1\ldots X_k}\|P_{X_1\ldots X_k})\Big), && (24)
\end{aligned}$$

which concludes the proof of the direct part of Theorem 1.

Converse. We next prove the converse part of Theorem 1, namely, that for any sequence of guessing strategies, $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}[G_k]$ is lower-bounded by the RHS of (3) (with $\rho = 1$). Our proof proceeds in two steps: first, we show that when $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ is equiprobable over a type class $\mathcal{T}^n(Q)$, every guessing strategy requires on average at least

$$2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})-\delta_n)} \tag{25}$$

guesses to recover $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$; second, by applying (25) in conjunction with the law of total expectation, we show that when $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k) \sim P^{\times n}_{X_1\ldots X_k}$, the RHS of (3) lower-bounds the exponential growth rate of the expected number of guesses of any guessing strategy. The first step is based on the following lemma, which we state without proof.

Lemma 1. *Let $\{i_1, \ldots, i_l\}$ be a size-l subset of $[1:k]$ and $\{i_{l+1}, \ldots, i_k\}$ its complement w.r.t. $[1:k]$. Suppose $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ is drawn equiprobably from a type class $\mathcal{T}^n(Q)$ and the l-tuple $(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_l})$ revealed to a guesser. Then, the expected number of guesses $\mathbb{E}_Q[G_{l+1} - G_l]$ until some component in the remaining $(k-l)$-tuple $(\boldsymbol{X}_{i_{l+1}}, \ldots, \boldsymbol{X}_{i_k})$ is revealed satisfies*

$$\begin{aligned}
&\min_{\mathcal{S}} \frac{1}{n} \log \mathbb{E}_Q[G_{l+1} - G_l] \\
&\quad \ge \min_{j\in[l+1:k]} \mathrm{H}_Q(X_{i_j} \mid X_{i_1}, \ldots, X_{i_l}) - \delta_n, && (26)
\end{aligned}$$

where $\{\delta_n\}$ is a positive sequence depending on $|\mathcal{X}_1 \times \cdots \times \mathcal{X}_k|$ and k only that decays to zero as n tends to infinity.

To apply Lemma 1, fix a permutation π^* on $[1:k]$ such that $\mathrm{H}_Q(X_{\pi^*(1)}) = \min_{i\in[1:k]} \mathrm{H}(Q_i)$ and such that for every $i \in [2:k]$,

$$\mathrm{H}_Q(X_{\pi^*(i)} \mid X_{\pi^*(1)}, \ldots, X_{\pi^*(i-1)}) = \min_{\substack{j\in\{1,\ldots,k\}\setminus \\ \{\pi^*(1),\ldots,\pi^*(i-1)\}}} \mathrm{H}_Q(X_j \mid X_{\pi^*(1)}, \ldots, X_{\pi^*(i-1)}). \quad (27)$$

That is, $\pi^*(1)$ equals i if X_i is the component of least entropy under Q; $\pi^*(2)$ equals i if X_i is the component of least conditional entropy given $X_{\pi^*(1)}$; and so forth. By Lemma 1, the expected number of guesses until the first affirmative answer $\mathbb{E}_Q[G_1]$ satisfies

$$\mathbb{E}_Q[G_1] \geq 2^{n(\mathrm{H}_Q(X_{\pi^*(1)})-\delta_n)}, \quad (28)$$

so the expected total number of guesses $\mathbb{E}_Q[G_k]$ satifies

$$\mathbb{E}_Q[G_k] \geq \mathbb{E}_Q[G_1] \geq 2^{n(\mathrm{H}_Q(X_{\pi^*(1)})-\delta_n)}. \quad (29)$$

We now argue that—starting from any strategy—(29) implies that without increasing its exponent, the guesser may employ a modified guessing scheme that first guesses $\boldsymbol{X}_{\pi^*(1)}$ followed by the original strategy with $\boldsymbol{X}_{\pi^*(1)}$ now assumed known. Indeed, the expected number of guesses of the modified scheme is larger than that of the original strategy by at most $|\mathcal{T}^n(Q_{\pi^*(1)})|$ and can therefore be upper-bounded by

$$\mathbb{E}_Q[G_k] + 2^{n\,\mathrm{H}_Q(X_{\pi^*(1)})}. \quad (30)$$

Because the exponential growth rate of a sum is dominated by that of the larger addend, (29) implies that

$$\liminf_{n\to\infty} \frac{1}{n} \log\Big(\mathbb{E}_Q[G_k] + 2^{n\,\mathrm{H}_Q(X_{\pi^*(1)})}\Big) = \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}_Q[G_k], \quad (31)$$

and we can thus assume without loss of optimality that the guesser first recovers $\boldsymbol{X}_{\pi^*(1)}$. With $\boldsymbol{X}_{\pi^*(1)}$ known, another application of Lemma 1 yields

$$\mathbb{E}_Q[G_k - G_1] \geq \mathbb{E}_Q[G_2 - G_1] \quad (32)$$
$$\geq 2^{n(\mathrm{H}_Q(X_{\pi^*(2)}|X_{\pi^*(1)})-\delta_n)}, \quad (33)$$

and as above, we conclude that without loss of optimality, the guesser recovers $\boldsymbol{X}_{\pi^*(2)}$ after $\boldsymbol{X}_{\pi^*(1)}$. Proceeding this way, we find that when $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ is equiprobable over $\mathcal{T}^n(Q)$, it is optimal (w.r.t. to minimizing the exponential growth rate of the expected total number of guesses) to guess its components in the order determined by π^*. Thus, for every guessing strategy,

$$\mathbb{E}_Q[G_k] \geq 2^{n(\max_i \mathrm{H}_Q(X_{\pi^*(i)}|X_{\pi^*(1)},\ldots,X_{\pi^*(i-1)})-\delta_n)}. \quad (34)$$

Recall from (19) that, when guessing the components of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ one-by-one in the order determined by a permutation π on $[1:k]$, the expected total number of guesses $\mathbb{E}_Q[G_k]$ satisfies

$$\mathbb{E}_Q[G_k] \leq 2^{n(\max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})+\delta_n)}. \quad (35)$$

In both (34) and (35) $\delta_n \downarrow 0$, so the two together imply:

Remark 1. *The permutation π^* minimizes*

$$\max_i \mathrm{H}_Q(X_{\pi(i)} \mid X_{\pi(1)}, \ldots, X_{\pi(i-1)}) \quad (36)$$

over all permutations $\pi\colon [1:k] \to [1:k]$.

Inequality (34) establishes the first step of the converse proof of Theorem 1. Returning to the actual setting with $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k) \sim P^{\times n}_{X_1,\ldots,X_k}$, we conclude the proof by taking the average over the events $\mathcal{A}(Q)$ of (12) and invoking (34):

$$\begin{aligned}
\mathbb{E}[G_k] &= \sum_Q \mathbb{E}_Q[G_k] \Pr[\mathcal{A}(Q)] && (37)\\
&\geq \sum_Q \Big(2^{n(\max_i \mathrm{H}_Q(X_{\pi^*(i)}|X_{\pi^*(1)},\ldots,X_{\pi^*(i-1)})-\delta_n)} \Pr[\mathcal{A}(Q)]\Big) && (38)\\
&= \sum_Q \Big(2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})-\delta_n)} \Pr[\mathcal{A}(Q)]\Big) && (39)\\
&\geq \sum_Q \Big(2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})-\delta_n)} 2^{-n(\mathrm{D}(Q\|P_{X_1\cdots X_k})+\delta'_n)}\Big) && (40)\\
&\geq \max_Q \Big(2^{n(\min_\pi \max_i \mathrm{H}_Q(X_{\pi(i)}|X_{\pi(1)},\ldots,X_{\pi(i-1)})-\delta_n)} 2^{-n(\mathrm{D}(Q\|P_{X_1\cdots X_k})+\delta'_n)}\Big), && (41)
\end{aligned}$$

where (39) is due to Remark 1; (40) is due to [9, Thm. 2.21]; and in (41) we have dropped all terms in the sum but the largest. Taking the limit and the logarithm on both sides of (41), and using the fact that the set of types is dense in set of all PMFs,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}[G_k] \geq \sup_{Q_{X_1\cdots X_k}} \Big(\min_\pi \max_i \mathrm{H}_Q\big(X_{\pi(i)} \mid X_{\pi(1)}, \ldots, X_{\pi(i-1)}\big) - \mathrm{D}(Q_{X_1\ldots X_k}\|P_{X_1\ldots X_k})\Big), \quad (42)$$

concluding the proof of the converse part of Theorem 1. ∎

III. Guessing With a Helper

In this section we prove Theorem 2. For lack of space and the similarity of the arguments to those in Section II, we only present an outline and restrict ourselves again to $\rho = 1$.

Achievability. To prove the direct part of Theorem 2, we will construct, for every sufficiently large n and every type $Q \in \mathcal{P}^n(\mathcal{X}_1 \times \cdots \times \mathcal{X}_k)$, a rate-$R$ helper φ_Q,

$$\varphi_Q\colon \quad \mathcal{T}^n(Q) \to \{0,1\}^{nR} \quad (43)$$
$$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k) \mapsto M_Q$$

and a helper-dependent guessing strategy $\mathcal{S}_Q(M_Q)$ satisfying

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{E}_Q[G_k(M_Q)]$$
$$\leq \inf_{Q_{U|X_1\dots X_k}:\ \mathrm{I}_{\tilde{Q}}(X_1,\dots,X_k;U)\leq R} \min_\pi \max_i \mathrm{H}_{\tilde{Q}}\left(X_{\pi(i)} \mid X_{\pi(1)},\dots,X_{\pi(i-1)},U\right), \quad (44)$$

where the entropy and mutual information on the RHS of (44) are computed w.r.t. $\tilde{Q} = Q \circ Q_{U|X_1\dots X_k}$, and where the argument $M_Q = \varphi_Q(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ in $\mathcal{S}_Q(M_Q)$ and $G_k(M_Q)$ emphasizes the dependence of the guessing strategy on the helper.

The pair $\left(\varphi_Q, \mathcal{S}_Q(M_Q)\right)$ is applied as follows: When $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k) \sim P^{\times n}_{X_1,\dots,X_k}$, the guesser is first revealed the empirical distribution $\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}$ of $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ and the helper's description $M \triangleq \varphi_{\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}}(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$, i.e., the result of applying the mapping (43) to $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ with $Q \leftarrow \hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}$. Given $\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}$ and M, the guesser follows the strategy $\mathcal{S}_{\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k}}(M)$ to recover $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$.

The direct part of Theorem 2 will follow from (44) by averaging over the events $\mathcal{A}(Q)$ (as in (20)–(23) of Section II.) It thus suffices to construct a helper φ_Q and a guessing strategy $\mathcal{S}_Q(M_Q)$ satisfying (44). To that end we shall need the Type Covering Lemma [9, Lemma 2.34]. It guarantees that for $R > \epsilon > 0$, a finite auxiliary alphabet $\mathcal{U}$, a sufficiently large n, and any type $\tilde{Q} \in \mathcal{P}^n(\mathcal{X}_1,\dots,\mathcal{X}_k,\mathcal{U})$ satisfying

$$\mathrm{I}_{\tilde{Q}}(X_1,\dots,X_k;U) \leq R - \epsilon, \quad (45)$$

the type class $\mathcal{T}^n(\tilde{Q}_{X_1,\dots,X_k})$ can be covered by 2^{nR} sequences from $\mathcal{T}^n(\tilde{Q}_U)$ in the sense that every $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)$ in $\mathcal{T}^n(\tilde{Q}_{X_1,\dots,X_k})$ is assigned some $\boldsymbol{u} \in \mathcal{T}^n(\tilde{Q}_U)$ such that the empirical distribution of $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{u})$ equals $\tilde{Q}$. We denote such a cover by $\mathcal{C}(\tilde{Q}) \subseteq \mathcal{T}^n(\tilde{Q}_U)$. We now construct a helper using the Type Covering Lemma as follows: We fix an auxiliary alphabet $\mathcal{U}$ and a small $\epsilon > 0$, choose a conditional type $Q^*_{U|X_1,\dots,X_k}$ that minimizes

$$\min_\pi \max_i \mathrm{H}_{\tilde{Q}}\left(X_{\pi(i)}|X_{\pi(1)},\dots,X_{\pi(i-1)},U\right) \quad (46)$$

over all $\tilde{Q} = Q \circ Q_{U|X_1,\dots,X_k}$ satisfying (45), and define

$$\tilde{Q}^* \triangleq Q \circ Q^*_{U|X_1\dots,X_k}. \quad (47)$$

The helper φ_Q describes to the guesser some $\boldsymbol{U} \in \mathcal{C}(\tilde{Q}^*)$ such that $\hat{Q}_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_k,\boldsymbol{U}} = \tilde{Q}^*$. Note that the Type Covering Lemma guarantees both the existence of $\boldsymbol{U}$ and the fact that nR bits suffice to describe it.

Based on the helper's description $\boldsymbol{U}$, we next construct the guessing strategy $\mathcal{S}_Q(M_Q)$ (where $M_Q = \boldsymbol{U}$). The construction is analogous to that of $\mathcal{S}^*(Q)$ in Section II: We choose a permutation π^* determined by $\tilde{Q}^*$ that minimizes

$$\max_i \mathrm{H}_{\tilde{Q}^*}\left(X_{\pi(i)} \mid X_{\pi(1)},\dots,X_{\pi(i-1)},U\right), \quad (48)$$

and the guesser recovers $\boldsymbol{X}_{\pi^*(1)},\dots,\boldsymbol{X}_{\pi^*(k)}$ one-by-one. Since $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k,\boldsymbol{U}) \in \mathcal{T}^n(\tilde{Q}^*)$,

$$\mathbb{E}_Q[G_i(M) - G_{i-1}(M)] \leq 2^{n\mathrm{H}_{\tilde{Q}^*}(X_{\pi^*(i)}|X_{\pi(1)},\dots,X_{\pi^*(i-1)},U)}, \quad (49)$$

and (44) follows from a chain of inequalities analogous to that in (15) to (18). Letting $\epsilon \downarrow 0$ concludes the proof of the direct part of Theorem 2. To prove the converse part of Theorem 2, we rely on the following lemma that we state without proof:

Lemma 2. *Let $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)$ be equiprobable over a type class $\mathcal{T}^n(Q)$. Given a guessing strategy, a rate-R helper φ, and a positive constant ϵ, define*

$$E_i \triangleq \frac{1}{n}\log \mathbb{E}_Q[G_i(M) - G_{i-1}(M)] + \epsilon, \quad i \in [1:k]. \quad (50)$$

There exists a positive decaying sequence $\{\delta_n\}$ (depending on $|\mathcal{X}_1 \times \dots \times \mathcal{X}_k|$ and k only), a permutation π on $[1:k]$, and k encoders,

$$\phi_i\colon \mathcal{X}_1^n \times \dots \times \mathcal{X}_k^n \to \{0,1\}^{nE_i}, \quad i \in [1:k], \quad (51)$$

with corresponding decoders,

$$\psi_i\colon \Big(\{0,1\}^{nE_i} \times \mathcal{X}^n_{\pi(i-1)} \times \dots \times \mathcal{X}^n_{\pi(1)} \times \{0,1\}^{nR}\Big) \to \mathcal{X}^n_{\pi(i)}, \quad i \in [1:k], \quad (52)$$

such that for all $i \in [1:k]$, with probability $1-\delta_n$

$$\psi_i\Big(\phi_i(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k), \boldsymbol{X}_{\pi(i-1)},\dots,\boldsymbol{X}_{\pi(1)}, \varphi(\boldsymbol{X}_1,\dots,\boldsymbol{X}_k)\Big) = \boldsymbol{X}_{\pi(i)}. \quad (53)$$

Using Lemma 2, one can show that

$$\frac{1}{n}\log \mathbb{E}_Q[G_k(M)]$$
$$\geq \max_i \frac{1}{n}\log \mathbb{E}_Q[G_i(M) - G_{i-1}(M)] \quad (54)$$
$$\geq \inf_{Q_{U|X_1\dots X_k}:\ \mathrm{I}_{\tilde{Q}}(X_1,\dots,X_k;U)\leq R} \min_\pi \max_i \mathrm{H}_{\tilde{Q}}\left(X_{\pi(i)} \mid X_{\pi(1)},\dots,X_{\pi(i-1)},U\right) - \delta'_n, \quad (55)$$

where $\tilde{Q} = Q \circ Q_{U|X_1,\dots,X_k}$. The converse part of Theorem 2 follows from (55) by averaging over the events $\mathcal{A}(Q)$ as in (37) to (41). ■

References

[1] M. M. Christiansen, K. R. Duffy, F. du Pin Calmon, and M. Médard, "Multi-user guesswork and brute force security," *IEEE Trans. Inf. Theory*, vol. 61, pp. 6876–6886, 2015.

[2] J. Massey, "Guessing and entropy," in *Proc. Int. Symp. Inf. Theory*, Jun. 1994, p. 204.

[3] E. Arıkan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, pp. 99 – 105, Jan. 1996.

[4] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, pp. 269 – 287, Jan. 2007.

[5] E. Arıkan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, pp. 1041 – 1056, May 1998.

[6] R. Graczyk and A. Lapidoth, "Variations on the guessing problem," in *Proc. Int. Symp. Inf. Theory*, Jun. 2018, pp. 231–235.

[7] N. Merhav and A. Cohen, "Universal randomized guessing with application to asynchronous decentralized brute–force attacks," *IEEE Trans. Inf. Theory*, vol. 66, pp. 114 – 129, Jan. 2020.

[8] R. Graczyk and A. Lapidoth, "Gray-Wyner and Slepian-Wolf guessing," in *Proc. Int. Symp. Inf. Theory*, Jun. 2020, pp. 2189–2193.

[9] S. Moser, *Advanced topics in information theory: lecture notes (ed. 4.5)*. ETH Zurich, Switzerland, 2020.